American Journal of Primatology 69:1–7 (2007)

BRIEF REPORT

Plant DNA Sequences from Feces: Potential Means for Assessing Diets of Wild Primates

BRENDA J. BRADLEY[1]*, MATHIAS STILLER[1], DIANE M. DORAN-SHEEHY[2], TARA HARRIS[1], COLIN A. CHAPMAN[3], LINDA VIGILANT[1], AND HENDRIK POINAR[4]
[1]*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[2]*Department of Anthropology, Stony Brook University, Stony Brook, New York*
[3]*Department of Anthropology and McGill School of Environment, McGill University, Montreal, Canada*
[4]*Department of Anthropology, McMaster University, Hamilton, Ontario, Canada*

Analyses of plant DNA in feces provides a promising, yet largely unexplored, means of documenting the diets of elusive primates. Here we demonstrate the promise and pitfalls of this approach using DNA extracted from fecal samples of wild western gorillas (*Gorilla gorilla*) and black and white colobus monkeys (*Colobus guereza*). From these DNA extracts we amplified, cloned, and sequenced small segments of chloroplast DNA (part of the *rbcL* gene) and plant nuclear DNA (*ITS-2*). The obtained sequences were compared to sequences generated from known plant samples and to those in GenBank to identify plant taxa in the feces. With further optimization, this method could provide a basic evaluation of minimum primate dietary diversity even when knowledge of local flora is limited. This approach may find application in studies characterizing the diets of poorly-known, unhabituated primate species or assaying consumer–resource relationships in an ecosystem. Am. J. Primatol. 69:1–7, 2007.

Key words: diet; DNA; feces; *rbcL*; *ITS-2*; DNA bar coding

INTRODUCTION

Knowledge of a species' diet is fundamental to understanding its place in a biological community and to structuring effective management plans for its conservation. Although researchers often obtain dietary information through direct observation, this is not an option under many circumstances, such as when the study subjects cannot be reliably observed or when food items are difficult to discern. In these cases, researchers rely on indirect methods of evaluating diet [Moreno-Black, 1978; van Wyk, 2000].

Contract grant sponsor: National Science Foundation; Contract grant number: SBR 9910399; Contract grant sponsor: LSB Leakey Foundation; Contract grant sponsor: Wenner Gren Foundation; Contract grant sponsor: Max Planck Society.

*Correspondence to: Brenda Bradley, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. E-mail: bjb37@cam.ac.uk

Received 30 March 2006; revised 28 July 2006; revision accepted 21 August 2006

DOI 10.1002/ajp.20384
Published online in Wiley InterScience (www.interscience.wiley.com).

Genetic analysis of fecal material [Höss et al., 1992] provides an alternative means of studying the diets of wild animals. By targeting plant and animal DNA segments whose sequences are highly variable and, in principle, species-specific, organisms can be identified by their ''DNA bar code'' [Moritz & Cicero, 2004]. Although DNA-based studies of predator diets are meeting with great success [Deagle et al., 2005; Jarman et al., 2004], molecular analyses of herbivore diets are proving much more difficult. Aside from a few molecular studies of fossilized sloth and human feces [Hofreiter et al., 2000; Poinar et al., 1998, 2001], DNA-based dietary analysis has not yet been employed to identify plant material. While animal matter can be readily identified by targeting mitochondrial DNA (mtDNA), for which there is a comprehensive GenBank reference collection, selection of target segment(s) for plant identification is not so straightforward. Choosing a target region involves tradeoffs among several factors: 1) minimizing target size to allow for amplification from the typically degraded DNA available in feces samples; 2) selection of regions of maximum sequence variability to allow for precise taxonomic identification; 3) design of primers that will amplify a wide variety of plant taxa while avoiding nonplant DNA; and 4) selection of segments for which there is an adequate reference collection. Here, we take a first step toward assessing and overcoming the difficulties of DNA-based analyses of primate diets by evaluating the efficacy of two potential markers for plant identification (a small segment of the *rbcL* gene and *ITS-2*).

METHODS

Fresh feces samples (n = 4 per species) were collected from wild, unhabituated western gorillas (*Gorilla gorilla*) at Mondika Research Center (Central African Republic and Republic of Congo) and from wild black and white colobus monkeys (*Colobus guereza*) at Kibale National Park (Uganda). For the gorilla samples, we obtained corresponding macroscopic data on diet [Doran et al., 2002]. For the four monkey samples, we had corresponding behavioral data from dawn-to-dusk focal animal follows spanning two consecutive days prior to defecation [Harris, 2005].

Fecal samples were desiccated and stored at ambient temperatures for up to 4 mo prior to extraction of total genomic DNA using the QIAGEN (Hilden, Germany) stool kit (following Bradley et al. [2001]).

For all samples, we amplified a 157-bp segment of the ribulose-bisphosphate carboxylase (*rbcL*) gene of the chloroplast genome following Poinar et al. [1998] using primers rbcLZ1: 5′-ATGTCACCACAAACAGAGACTAAAGCAAGT-3′ and rbcL19b: 5′-CTTCTTCAGGTGGAACTCCAG -3′. The four monkey DNA samples were also amplified at the ~350-bp second internal transcribed spacer of the nuclear ribosomal genes (*ITS-2*) using primers rD5-ITS2: 5′-TCCTCCGCTTATTGATATGC-3′ and rb1-ITS2f 5′-CGATACTTGGTGTGAATTGCAG-3′. PCR amplification was carried out in a total volume of 30 μl consisting of 5 μl DNA template (minimum of 225 pg total DNA), 2 mM $MgCl_2$, 30 mg bovine serum albumin (BSA), 250 μM each dNTPs (nucleotides), 200 nM each primer, five Units Amplitaq Gold and 1 × polymerase chain reaction (PCR) buffer (Perkin-Elmer; Rodgau, Germany). PCR conditions were as follows: 94°C denaturation for 5 min, 40 cycles of 92°C for 15 sec, 57°C (*rbcL*) or 59°C (*ITS-2*) for 1 min, and 72°C for 1 min, with a final extension at 72°C for 10 min. Three amplification products per sample were cloned directly into TA cloning vectors (Invitrogen; Karlsruhe, Germany) following the manufacturer's instructions. Colony PCR was performed according to Kilger et al. [1997] followed by PCR cleanup and cycle sequencing

(Big-Dye^{TM}, Applied Biosystems; Darmstadt, Germany) as described in Ebersberger et al. [2002]. Cycle-sequencing reaction fragments were separated and visualized on an ABI3700 automated sequencer (Applied Biosystems).

Sequences were aligned by eye and clustered into groups of identical sequences. A consensus sequence was generated from each cluster of clones. As is customary, sporadic substitutions found only in products from a single amplification were attributed to misincorporations by the *Taq* polymerase or to DNA damage and were excluded [Poinar et al., 2001]. Taxonomic affiliations of consensus sequences were compared to the plant sequences at GenBank using the program BLASTN (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) [Altschul et al., 1997]. A consensus sequence was assigned to a taxon when the sequence exclusively matched ($<1.0\%$ mismatch) members of that taxon in the database to the exclusion of all other taxa (following Poinar et al. [1998]).

To have a representative reference collection of sequences from plant taxa consumed by the colobus monkeys, we obtained dried samples from seven plant species that comprised the majority of the monkey diet during the sampling period [Harris, 2005]. DNA from these samples was extracted (following Gustincich et al. [1991]), and then amplified and directly sequenced (as above) at both *rbcL* and *ITS-2*. Sequences were deposited in GenBank as: *Celtis africana* (AY702566, AY702559); *Celtis durandii* (AY702561, AY702554); *Albizia grandibracteata* (AY702565, AY702558); *Markhamia lutea* (formerly *platycalyx*) (AY702564, AY702557); *Strombosia scheffleri* (AY702560, AY702553); *Spathodea campanulata* (AY702562, AY702555); and *Premna angolensis* (AY702563, AY702556).

RESULTS

A total (excluding singlets) of 255 *rbcL* clones were sequenced from the four gorilla feces (range 52–78 per sample) and a total of 299 *rbcL* clones were sequenced from the four monkey feces (range 47–101). In addition, 308 *ITS-2* clones were sequenced from the four monkey samples (range 61–90 clones). The number of clones per consensus sequence/identification varied from two to 70. In some cases, two *rbcL* consensus sequences were assigned to the same plant order or family. In these cases one sequence exactly matched the reference sequence(s) while the other sequence mismatched the reference sequence(s) at one nucleotide. These sequences are assumed to represent a minimum of two different plant species belonging to the same family.

The total numbers of dietary items detected per sample using each method are shown in Table I. By examining *rbcL* sequences we identified a minimum of 16 different plant items in the four gorilla feces (three to the level of subclass, five to the level of order, and eight to the level of family), with the number of items per sample ranging from 5 to 8. By examining *rbcL* sequences in the monkey feces we detected four different plant families (2–4 per sample), one of which could be identified more precisely to the species level by analysis of *ITS-2* sequences in the feces.

Table II shows the specific plant taxa identified in the eight feces samples using each method. The plant taxa expected in the monkey feces based on behavioral observation of feeding behavior generally corresponded to those plant taxa identified by genetic analyses. All four plant families on which the monkeys were observed to feed were identified by their DNA (*rbcL*) sequences in the feces. However, the family Bignoniaceae was identified in samples I and III by behavioral observation and samples II and IV by DNA analyses. The family Salicaceae was genetically identified in all four monkey feces, but none of the

TABLE I. Total Number of Plant Items Identified by Each Method and the Corresponding Levels of Taxonomic Precision*

	Gorillas								Colobus monkeys							
	rbcL				macro				rbcl, ITS-2				obs			
Taxonomic level	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Order	4	4	3	4												
Family	3	1	5	3					3	4	2	4				
Species					1	3	3	1		1	1	1	3	3	4	3
Total	7	5	8	7	1	3	3	1	3	5	3	5	3	3	4	3

*Roman numerals represent the four matched samples from each taxa.
Macro, identified through macro-analyses of feces (gorilla samples only); Obs, identified by direct observations of feeding (monkey samples only); rbcL, identified by *rbcL* sequences (gorilla and monkey samples); ITS-2, identified by *ITS-2* sequences (monkey samples only).

known plants on which the monkeys feed (during any season) are in this family. The *ITS-2* sequences allowed plant identification to the species level, and one (*Celtis africana*) of the four plant species observed to be eaten by the monkeys was also correctly identified in the corresponding feces samples (II, III, IV). However, a large number of the *ITS-2* sequences from the feces could not be assigned. For about 10% of the sequences (28 of 308 sequences; five of 13 consensuses) the closest match(es) in GenBank were to sequences from fungus (e.g., GenBank AF413092; categorized as ''unknown-fungi'' in Table II), but even these matched by only about 60%. More surprisingly, for almost two-thirds of the sequences (191 of 308 sequences; five of 13 consensuses) there were no sequences in GenBank that matched within 50% (categorized as ''unknown'' in Table II). This suggests that the *ITS-2* primers are amplifying an additional unknown plant genome locus or loci, even though when the *ITS-2* primers were subjected to a BLASTN search, the corresponding matches were only from *ITS-2* sequences.

DISCUSSION

We genetically identified a minimum of five to eight different plant items in the gorilla feces and three to five items in the monkey feces. We had no prior knowledge of the items consumed by the individual gorillas in the days preceding sample collection, but 15 of the 16 plant taxa identified in the gorilla feces are known to occur in this region [Harris, 2002] and at least one species from each of these taxa is eaten by gorillas [Rogers et al., 2004]. The few plant taxa (one to three species per sample) identified by macroanalyses (as described in Doran et al. [2002]) were not the same as those identified by DNA analysis of the same gorilla feces (Table II). This is probably because the molecular analyses of the gorilla feces targeted only the chloroplast marker (*rbcL*) and may have preferentially amplify DNA from chloroplast-rich tissues, such as leaves or stems, whereas macroanalysis usually identifies fruit and seed fragments. Thus, the *rbcL* marker might be more appropriate for studies of folivores than studies of frugivores.

The plant families expected in the monkey feces based on direct focal-animal observations of foraging behavior during the 2 days prior to sample collection [Harris, 2005], generally corresponded to those plant taxa identified by genetic analyses of *rbcL* (Table II). All four plant families on which the monkeys were

TABLE II. Plant Taxa Identified in the Diets of Wild Western Gorillas and Black and White Colobus Monkeys Using Various Methods*

Subclass Order Family Genus species	Gorillas				Colobus monkeys			
	I	II	III	IV	I	II	III	IV
Asteridae								
Lamiale								
Gesneriaceae	rbcL							
Lamiacea					obs, rbcL	obs, rbcL	obs, rbcL	obs, rbcL
Premna angolensis					obs	obs	obs	obs
Bignoniaceae					obs	rbcL	obs	rbcL
Markhamia lutea					obs		obs	
Sapotaceae (macro only)								
Chrysophyllum lacoutiana				macro				
Unspecified	rcbL	rcbL		rbcL				
Gentianales								
Unspecified			rbcL					
Commelinidae								
Lilliale								
Arecaceae (= Palmae)	rbcL		rbcL					
Unspecified		rbcL						
Zingiberales								
Unspecified	rbcL	rbcL (2)	rbcL (2)	rbcL (2)				
Rosidae								
Fabales								
Fabaceae (= Leguminosae)	rbcL	rbcL	rbcL (2)	rbcL				
Celastrales								
Celastraceae			rbcL (2)					
Rosales								
Moraceae				rbcL				
Ulmacea					obs, rbcL	obs, rbcL	obs, rbcL	obs, rbcL
Celtis durandii					obs	obs	obs	obs
Celtis africana						obs, ITS	obs, ITS	obs, ITS
Unspecified				rbcL				
Sapindales								
Sapindaceae (macro only)								
Pancovia laurentii			macro					
Nitrariaceae				rbcL				
Irvingiaceae (macro only)								
Klainedoxa gabonensis	macro	macro	macro					
Malpighale								
Salicaceae					rbcL	rbcL	rbcL	rbcL
Malvale								
Tiliaceae (macro only)								
Duboscia macrocarpa		macro	macro					
Grewia oligoneura		macro						
Unspecified	rbcL (2)							
Unknown-fungi							ITS	ITS
Unknown					ITS	ITS	ITS	ITS

*Roman numerals represent the four matched samples from each taxa.
(2), two consensus sequences were identified to the same family and thought to represent 2 different species from that plant family; Macro, identified through macro-analyses of feces (gorilla samples only); Obs, identified by direct observations of feeding (monkey samples only); rbcL, identified by *rbcL* sequences (all samples); ITS, identified by *ITS-2* sequences (monkey samples only).

observed to feed were identified by their DNA (*rbcL*) sequences in the feces. Inexplicably, the family Bignoniaceae was identified in samples I and III by behavioral observation and samples II and IV by DNA analyses. The sequences from the family Salicaceae, which is not know to occur in the colobus monkey diet but was genetically identified in the monkey feces, might derive from the many vines in the monkey diet that have not been taxonomically classified.

Although the segment of *rbcL* sequenced here is highly variable, *rbcL* identifications are necessarily limited by the size of the target DNA segment (157 bp), which, at this locus, allows taxonomic classification to the level of family or order [Poinar et al., 1998]. Since the fragment amplified here was originally selected for ancient DNA analyses, it is necessarily small in order to allow for ready amplification of degraded DNA. However, larger target fragments of *rbcL* might be amplified from fresh feces samples, which would capture a greater amount of genetic variability and thereby improve precision.

The *ITS-2* sequences allowed plant identification to the species level, and one (*Celtis africana*) of the four plant species observed to be eaten by the monkeys was correctly identified in the corresponding feces samples (II, III, IV). However, a large number of the *ITS-2* sequences from the feces could not yet be assigned. We suggest that *ITS-2* holds great promise as a marker for identifying plant species, but primers need to be designed to specifically target plant groups of interest (as in Jarman et al. [2004] regarding prey items). Although the reference collection for *ITS-2* sequences is currently less than that for *rbcL*, the database could be easily expanded by specifically sequencing herbarium specimens of potential interest, as was done here.

With further optimization, this approach should prove especially valuable for those struggling to obtain feeding data on elusive primates. Many researchers are already collecting fecal samples from study subjects for other types of DNA analyses, and adding a DNA-based dietary component to an on-going field project would be relatively inexpensive in terms of material (<0.10 g of fecal material needed per extraction), time and money (10 samples plus controls could be analyzed in 2–4 weeks for approximately $500 worth of consumables).

Ideally, researchers hoping to understand the ecology of elusive primates should combine multiple approaches to studying diet, including examining feeding remains, studying behavior when possible, and conducting macro-, chemical, and DNA-based analyses of feces [Ortmann et al., 2006].

ACKNOWLEDGMENTS

We thank the field assistants at Mondika and Kibale, and we thank M. Kuch, R. Blatter, and H. Siedel for laboratory assistance. We especially thank D. Harris for information on African flora.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Bradley BJ, Chambers KE, Vigilant L. 2001. Accurate DNA-based sex identification of apes using non-invasive samples. Conserv Genet 2:179–181.

Deagle B, Tollit DJ, Jarman SN, Hindell MA, Trites AW, Gales NJ. 2005. Molecular scatology as a tool to study diet: analysis of prey in scats from captive Steller sea lions. Mol Ecol 14:1831–1842.

Doran DM, McNeilage A, Greer D, Bocian C, Mehlman P, Shah N. 2002. Western lowland gorilla diet and resource availability: new evidence, cross-site comparisons, and reflec-

tions on indirect sampling methods. Am J Primatol 58:91–116.

Ebersberger I, Metzler D, Schwarz C, Pääbo S. 2002. Genome-wide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70:1490–1497.

Gustincich S, Manfioletti G, Del Sal G, Schneider C, Carninci P. 1991. A fast method for high-quality genomic DNA extraction from whole human blood. Biotechniques 11:298–302.

Harris DJ. 2002. The vascular plants of the Dzanga-Sangha Reserve, Central African Republic. Scripta Botanica Belgica, Vol. 23. Meise, Belgium: National Botanic Garden of Belgium (NBGB). 274p.

Harris TR. 2005. Roaring, intergroup aggression, and feeding competition in black and white colobus monkeys (*Colobus guereza*) at Kanyawara, Kibale National Park, Uganda. PhD Thesis, Yale University, New Haven, CT. 253p.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. 2000. A molecular analysis of ground sloth diet through the last glaciation. Mol Ecol 9: 1975–1974.

Höss M, Kohn M, Pääbo S. 1992. Excrement analysis by PCR. Nature 359:99.

Jarman S, Deagle BE, Gales NJ. 2004. Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. Mol Ecol 13: 1313–1322.

Kilger C, Krings M, Poinar H, Pääbo S. 1997. ''Colony sequencing'': direct sequencing of plasmid DNA from bacterial colonies. Biotechniques 22:412–414, 416, 418.

Moreno-Black G. 1978. The use of scat samples in primate diet analysis. Primates 19: 215–221.

Moritz C, Cicero C. 2004. DNA barcoding: promise and pitfalls. PLoS 2:1529–1531.

Ortmann S, Bradley B, Stolter C, Ganzhorn J. 2006. Estimating the quality and composition of wild animal diets—a critical survey of methods. In: Hohmann G, Robbins M, Boesch C, editors. Feeding ecology in apes and other primates. Cambridge: Cambridge University Press. p 395–418.

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, Pääbo S. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. Science 281:402–406.

Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, Kuder T, Spaulding WG, Bryant VM, Cooper A, Pääbo S. 2001. A molecular analysis of dietary diversity for three archaic Native Americans. Proc Natl Acad Sci USA 98:4317–4322.

Rogers ME, Abernethy K, Bermejo M, Cipolletta C, McFarland K, Nishihara T, Remis M, Tutin CEG. 2004. Western gorilla diet: a synthesis from six sites. Am J Primatol 64: 173–192.

van Wyk JH. 2000 Seasonal variation in stomach contents and diet composition in the large girdled lizard, *Cordylus giganteus* (Reptilia: Cordylidae) in the Highveld grasslands of the northeastern Free State, South Africa. Afr Zool 35:9–27.